CHEMBIOCHEM

DOI: 10.1002/cbic.200700379

Processing of N-Terminal Unnatural Amino Acids in Recombinant Human Interferon-β in *Escherichia coli*

Aijun Wang,*[a] Natalie Winblade Nairn,[a] Richard S. Johnson,[b] David A. Tirrell,[c] and Kenneth Grabstein[a]

Incorporation of unnatural amino acids into recombinant proteins represents a powerful tool for protein engineering and protein therapeutic development. While the processing of the N-terminal methionine (Met) residues in proteins is well studied, the processing of unnatural amino acids used for replacing the N-terminal Met remains largely unknown. Here we report the effects of the penultimate residue (the residue after the initiator Met) on the processing of two unnatural amino acids, L-azidohomoalanine (AHA) and L-homopropargylglycine (HPG), at the N terminus of recombinant human interferon-β in E. coli*. We have identified specific amino acids at the penultimate position that can be used to efficiently retain or remove N-terminal AHA or HPG. Retention of N-terminal AHA or HPG can be achieved by choosing amino acids with large side chains (such as Gln, Glu, and His) at the penultimate position, while Ala can be selected for the removal of N-terminal AHA or HPG. Incomplete processing of N-terminal AHA and HPG (in terms of both deformylation and cleavage) was observed with Gly or Ser at the penultimate position.*

Introduction

With the rare exceptions of selenocysteine[1,2] and pyrrolysine,[3,4] proteins from all known organisms are made from 20 amino acid building blocks. The limited number of proteinogenic amino acids stands in contrast to the much larger repertoire of unnatural amino acids (UAAs) that have been shown to exhibit translational activity.[5] Introduction of UAAs into proteins creates new opportunities for protein structural and functional studies, for protein engineering, and for protein therapeutic development. With the unique properties of UAAs, proteins can be designed with increased thermal stability,[6–8] desired biophysical properties, and novel functions.[9–13] In addition, proteins can be modified site-specifically at the unique functional groups on UAAs to generate protein conjugates with enhanced therapeutic efficacies or improved pharmacokinetic profiles.[14,15]

Several different strategies have been developed to introduce UAAs into recombinant proteins in both prokaryotic and eukaryotic expression systems.[16–23] UAAs can be incorporated either "residue-specifically", by replacing all copies of a particular amino acid in the protein sequence, or "site-specifically" at programmed position(s). The site-specific method requires introduction into the expression host of a modified aminoacyl-tRNA synthetase and cognate tRNA. In contrast, the residue-specific method exploits the promiscuity of the aminoacyl-tRNA synthetases (wild type or mutant) toward certain UAAs that are structurally similar to the natural amino acids. The residue-specific method has been used to incorporate several Met analogues, including L-azidohomoalanine (AHA) and L-homopropargylglycine (HPG; Scheme 1), into recombinant proteins, replacing Met at both N-terminal and internal sites.[24–27] The reactive side chains of AHA or HPG enable specific chemical modification of recombinant proteins without interference from reactions with any of the natural amino acids.[25,28–30] The simplicity of the residue-specific method and its high protein yield, make it an attractive approach for many protein engineering needs.

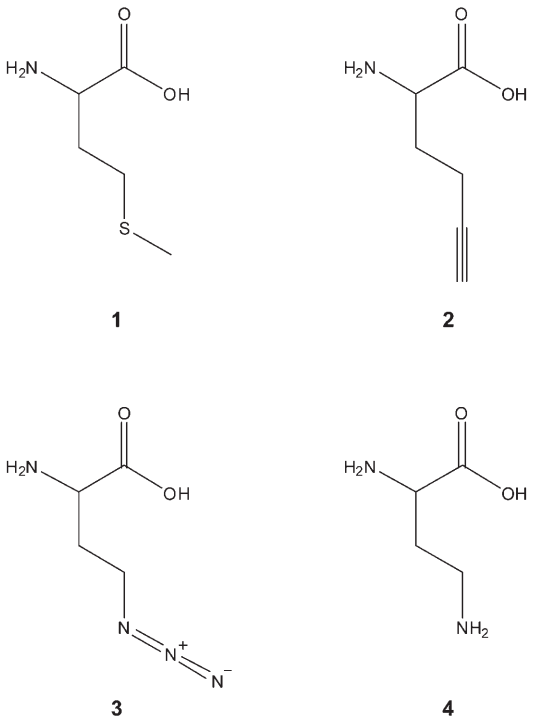
Whenever the Met-substitution method is used, the N-terminal Met will be replaced with the analogue. This may or may not be desired depending on the protein and the preferred site(s) of modification. Therefore, it is important to define whether the rules for N-terminal methionine excision apply to UAAs. It is known that in prokaryotes, protein translation generally initiates with *N*-formylmethionine (f-Met). The f-Met on the nascent polypeptide is post-translationally deformylated by peptide deformylase.[31,32] Furthermore, in more than half of the *E. coli* proteins, the N-terminal Met is subsequently removed by methionine aminopeptidase (MetAP).[33,34] It is well established that the specificity of MetAP depends primarily on the identity of the second or penultimate residue of the protein in *E. coli*.[34–37] While the presence of UAAs at the N terminus of recombinant proteins has been observed,[27,38] a systematic study on the processing of UAAs at the N terminus has

[a] *Dr. A. Wang, Dr. N. Winblade Nairn, Dr. K. Grabstein*
Allozyne Inc., 1616 Eastlake Ave E.
Seattle WA 98102 (USA)
Fax: (+1) 206-957-7399
E-mail: awang@allozyne.com

[b] *Dr. R. S. Johnson*
Homestead Corporation, 1616 Eastlake Ave E.
Seattle, WA 98102 (USA)

[c] *Prof. D. A. Tirrell*
Division of Chemistry and Chemical Engineering
Joseph J. Jacobs Institute for Molecular Engineering for Medicine
California Institute of Technology
Pasadena, CA 91125 (USA)

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.



Scheme 1. Chemical structures of Met (**1**), HPG (**2**), AHA (**3**), and DAB (**4**).

not been reported. Using a modified human interferon-β (IFNβ) protein, we have identified specific amino acids that, when present at the penultimate site, enable controlled retention or removal of N-terminal AHA or HPG.

Results and Discussion

Gene construction and recombinant protein expression with AHA or HPG

A synthetic gene with optimal *E. coli* codon usage was synthesized by using PCR with overlapping oligonucleotides. The synthetic gene encodes a modified human IFNβ (~20 kDa) with only one Met codon, which is located at the N terminus. Different penultimate residues were introduced to the protein by site-directed mutagenesis (Table 1). This system provided a tool to study the effect of the penultimate residue on the processing of the Met surrogate, AHA or HPG, at the N terminus. The synthetic gene was cloned into the pQE30 expression vector under control of the inducible promoter T5-lac-lac. Upon addition of IPTG in AHA- or HPG-supplemented media in a Met auxotrophic *E. coli* strain, abundant IFNβ was induced. Figure 1 shows a typical expression profile of IFNβ-2E with AHA or HPG, along with control samples expressed in media

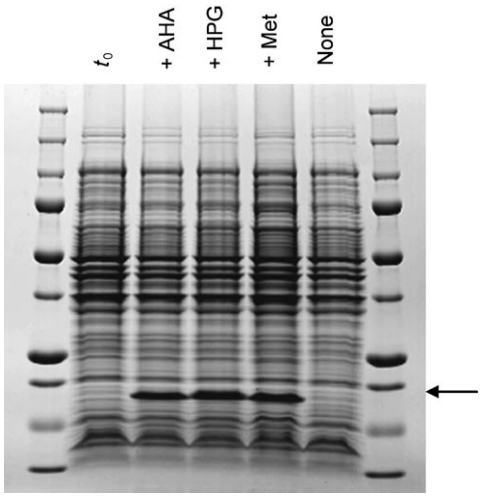


Figure 1. SDS-PAGE (4–20%) of recombinant IFNβ-2E expression profile. *E. coli* cell lysates prior to induction (t_0), and 2 h post-IPTG induction in Met-free medium supplemented with AHA, HPG, Met or no Met or analogue, were analyzed and protein bands were visualized by Coomassie blue staining. The protein markers in lanes 1 and 7 are 250, 150, 100, 75, 50, 37, 25, 20, 15, and 10 kDa. The induced IFNβ protein band is indicated by the arrow.

with and without Met. Recombinant protein was expressed to the same level with AHA or HPG as with Met; background expression (in the absence of Met) was very low, and basal expression (prior to induction) was absent. Similar expression profiles were observed for all IFNβ variants listed in Table 1, whether they were expressed with AHA or with HPG. Our results agree well with previous observations that AHA and HPG support high protein expression yields and can be incorporated efficiently into recombinant proteins.[25,39,40]

MALDI-MS analysis for penultimate residues that favor retention of N-terminal AHA or HPG

MetAP catalyzes the removal of N-terminal Met from cellular proteins. However, not all proteins are susceptible to this

Table 1. N-terminal oligonucleotide and peptide sequences, and calculated M_W for the N-terminal peptides. The penultimate amino acids and corresponding codons are highlighted.

Name	N-terminal oligonucleotide sequence	peptide sequence	calculated peptide M_W [Da] X = Met	X = AHA	X = HPG	X cleaved
IFNβ-2A	atg**gcg**tataatctgttaggctttctgcaacgt	X**A**YNLLGFLQR	1324.70	1319.62	1302.63	1193.66
IFNβ-2S	atg**agc**tataatctgttaggctttctgcaacgt	X**S**YNLLGFLQR	1340.69	1335.61	1318.62	1209.65
IFNβ-2G	atg**ggc**tataatctgttaggctttctgcaacgt	X**G**YNLLGFLQR	1310.68	1305.60	1288.61	1179.64
IFNβ-2H	atg**cac**tataatctgttaggctttctgcaacgt	X**H**YNLLGFLQR	1390.72	1385.64	1368.65	1259.68
IFNβ-2Q	atg**cag**tataatctgttaggctttctgcaacgt	X**Q**YNLLGFLQR	1381.72	1376.64	1359.65	1250.68
IFNβ-2E	atg**gag**tataatctgttaggctttctgcaacgt	X**E**YNLLGFLQR	1382.70	1377.62	1360.63	1251.66

cleavage. While the antepenultimate (third) residue might also influence N-terminal residue excision, the substrate specificity of MetAP is primarily determined by the side-chain length of the penultimate residue in a given protein.[34–37] Proteins with small penultimate residues are preferred substrates for MetAP. The extent of cleavage by MetAP decreases as the size of the penultimate residue increases. Nine amino acids with large side chains (namely Gln, His, Glu, Phe, Lys, Tyr, Trp, Arg, and Met) have been reported to render proteins completely resistant to processing by MetAP.[35,36] When these residues occupy the penultimate position, the N-terminal Met remains on the protein. The observed substrate specificity of MetAP led to the hypothesis that the substrate-binding pocket in MetAP has two sites: one site specific for N-terminal Met, and another site with limited space to accommodate the penultimate residue.[36] This substrate binding model is supported by crystallographic structure studies of *E. coli* MetAP and its substrate-like inhibitor complex,[41,42] and by mutagenesis studies in which engineered MetAP enzymes with enlarged substrate binding pockets were able to remove N-terminal Met from proteins with bulky penultimate residues.[35]

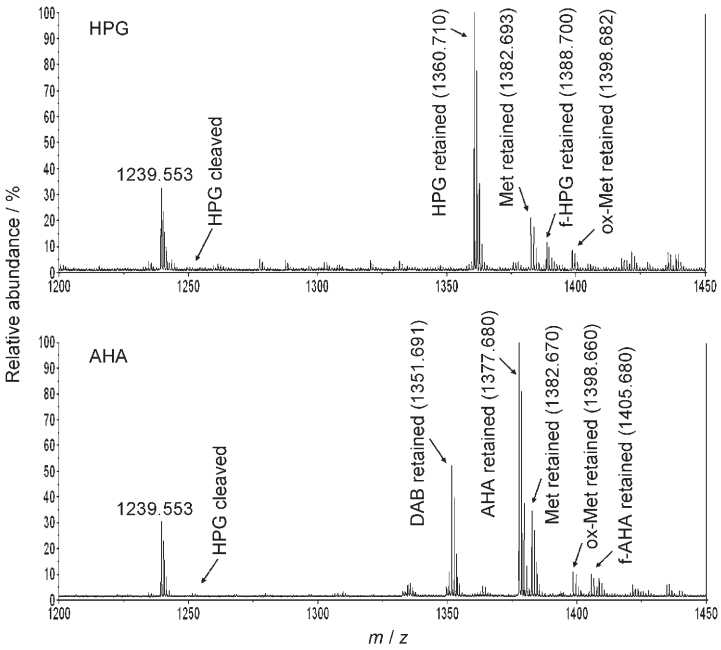


Figure 2. MALDI mass spectra of the tryptic peptides for IFNβ-2Q expressed with HPG and AHA (Table 1). The N-terminal peptide variants are labeled, and the *m/z* values of the protonated molecules are in parentheses. The signal at 1239.5 is assigned to an internal IFNβ peptide.

Based on these studies, we speculated that the nine large amino acids would also protect N-terminal UAAs from cleavage by MetAP. To test our hypothesis, we generated three proteins (IFNβ-2Q, IFNβ-2E, and IFNβ-2H) with Gln, Glu, and His as the penultimate residues, respectively. The antepenultimate residue Tyr in IFNβ was kept the same and the effects of the antepenultimate residue were not investigated in the current study. Each protein was expressed with AHA or HPG, separated by SDS-PAGE, digested by trypsin, and analyzed by MALDI-MS. MALDI-MS analysis was used to verify the integrity of these two UAAs in the recombinant proteins, and to examine the potential processing of these two UAAs as the N-terminal residue in *E. coli*, including deformylation by peptide deformylase and cleavage by MetAP. As controls, each protein was also expressed with Met and analyzed by MALDI-MS, and the results demonstrated complete retention of N-terminal Met, as expected (data not shown).

MALDI-MS analysis of all three proteins expressed with AHA or HPG showed that N-terminal AHA and HPG are effectively retained. Figure 2 shows MALDI-MS spectra of IFNβ-2Q as an example. It should be mentioned that of all the tryptic peptides, the N-terminal peptide was the most efficiently ionized; furthermore, the identities of the major peaks were confirmed by MS/MS analysis. When IFNβ-2Q was expressed with HPG, the most abundant ion ($[M+H]^+$ *m/z* 1360.7) was identified as the intact HPG-containing peptide (Table 1). This result indicates that HPG was incorporated efficiently into the recombinant protein, and remained stable in vivo. Virtually no cleaved peptide was detected at the expected $[M+H]^+$ *m/z* of 1251.7.

Low abundance ions from Met-containing peptides (Met and oxidized-Met) were also detected by MALDI-MS, and agreed well with the low background expression observed in Met-free medium (Figure 1). The small amounts of Met-containing species might result from incomplete depletion of Met from the induction media, from the intracellular pool of amino acids, and/or from turnover of endogenous *E. coli* proteins during induction. Retention of Met in these low-abundance peptides is fully expected, as the penultimate Gln residue inhibits cleavage by MetAP.

A small amount of the N-terminal formylated HPG (f-HPG) peptide was detected at $[M+H]^+$ *m/z* 1388.7, as compared to the large peak of deformylated HPG peptide at 1360.7. Apparently, deformylation of N-terminal f-HPG by *E. coli* peptide deformylase is efficient, yet incomplete. The incomplete deformylation observed here might not be unique to HPG or other UAAs, as it has also been observed for recombinant proteins expressed at high levels in the absence of UAAs.[43,44] Increasing peptide deformylase activity in the expression host has been shown to yield completely deformylated recombinant proteins.[43,44] This strategy could also find application for engineered proteins with UAAs if complete deformylation is required.

Similar results were obtained for IFNβ-2Q expressed with AHA (Figure 2). The only difference (as compared to HPG) was the appearance of a new peak at $[M+H]^+$ *m/z* 1351.7, which was identified as the peptide that bore N-terminal 2,4-diaminobutyric acid (DAB, Scheme 1). The occurrence of DAB signals in MALDI-MS analysis of AHA-containing peptides had been observed previously.[25] Since azides are known to be photosensitive,[45,46] we believe that the observed DAB signal is the reduction product of AHA due to laser irradiation of the MALDI-MS analysis. This interpretation is supported by the lack of a DAB peak in N-terminal sequence (NTS) analysis of AHA-peptide,[25]

and by our own NTS analysis discussed later. The retention of AHA and HPG at the N terminus of IFNβ also correlated well with previous observations in which other UAAs were found to replace Met at the N terminus of proteins.[27,38]

The presence of HPG and AHA at the N terminus of IFNβ implies that not only can these UAAs be charged to the initiator tRNA^{fMet} by the catalytic promiscuity of the methionyl-tRNA synthetase, but also the charged UAA-tRNA^{fMet} molecules can be further formylated by the methionyl-tRNA^{fMet} formyltransferase (MTF) and recognized by the initiation factor to start translation in *E. coli*. The formylation of UAA-tRNA^{fMet} by MTF is not surprising since the substrate specificity of MTF is primarily determined by a set of nucleotides in the tRNA^{fMet} molecule.[47–49] Post-translationally, the formylated UAA at the N terminus of nascent polypeptide can be deformylated by the peptide deformylase, as shown in Figure 2. Deformylated UAA at the N terminus can be further cleaved by MetAP if an appropriate penultimate residue is present, as demonstrated in the next section.

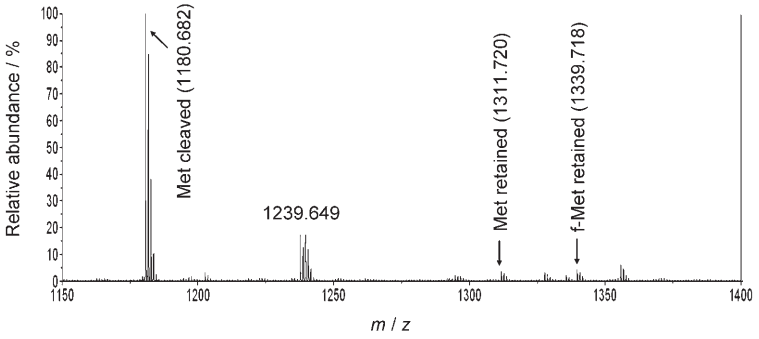


Figure 3. MALDI mass spectra of the tryptic peptides for IFNβ-2G expressed with Met (Table 1). The N-terminal peptide variants are labeled, and the *m/z* values of the protonated molecules are in parentheses. The signal at 1239.6 is assigned to an internal IFNβ peptide.

MALDI-MS analysis for penultimate residues that favor removal of N-terminal AHA or HPG

The two smallest amino acids, Ala and Gly, have been reported to be the most efficient penultimate residues for N-terminal cleavage of recombinant proteins by MetAP; Pro, Ser, Thr, and Val were reported to be slightly less efficient.[35] However, proteomic studies of 223 unique genes in *E. coli* indicated that all N-terminal Met residues are cleaved when the penultimate residue is either Ser or Ala; N-terminal Met was cleaved to variable extents when the penultimate residue was Gly, Thr, or Pro.[33] Based on these results, we selected three candidates, Ala, Ser, and Gly, for controlled removal of N-terminal AHA and HPG. With any of these three amino acids at the penultimate position, the N-terminal Met was efficiently removed in IFNβ expressed with Met (Figure 3 shows an example). The MALDI-MS spectra for the three proteins (IFNβ-2A, IFNβ-2S, and IFNβ-2G) expressed with HPG are shown in Figure 4.

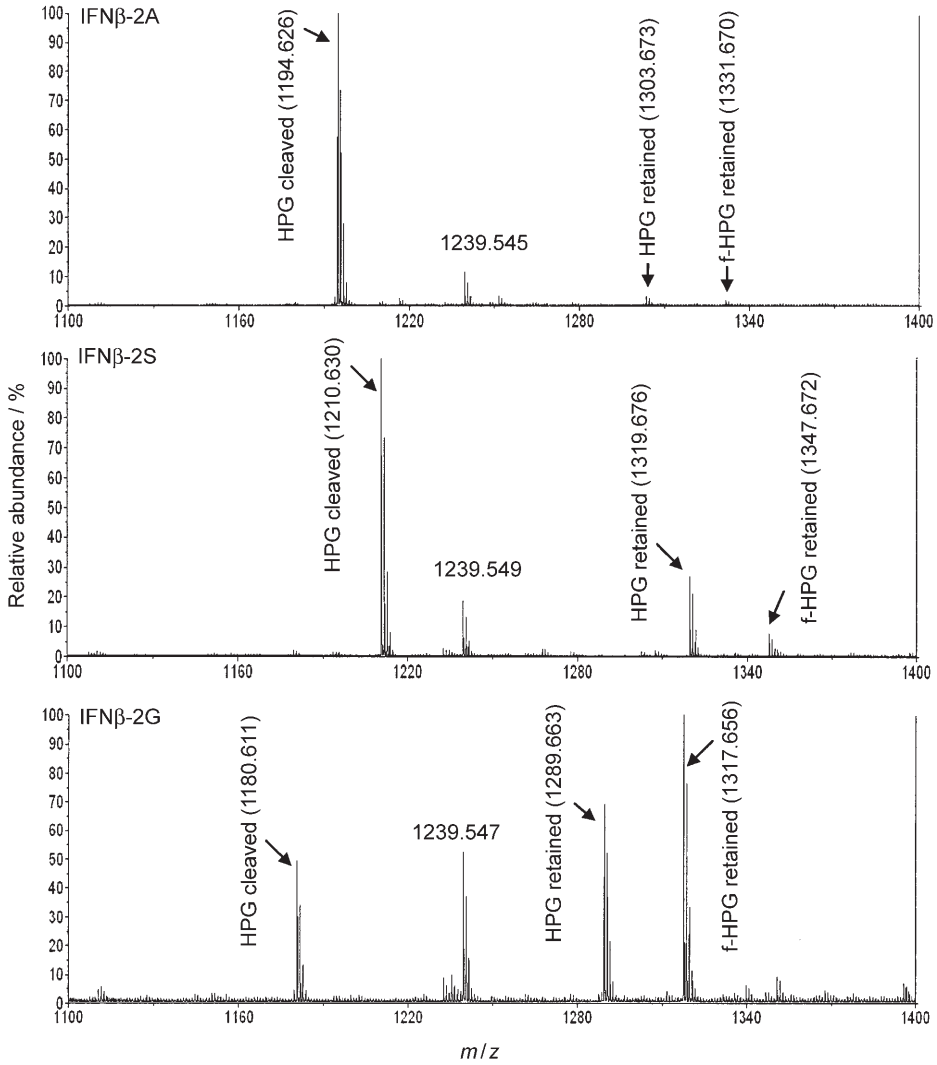


Figure 4. MALDI mass spectra of the tryptic peptides for IFNβ-2A, IFNβ-2S, and IFNβ-2G proteins expressed with HPG (Table 1). The N-terminal peptide variants are labeled, and the *m/z* values of the protonated molecules are in parentheses. The signal at 1239.5 is assigned to an internal IFNβ peptide.

For IFNβ-2A, the most abundant ion ($[M+H]^+$ *m/z* 1194.7) was identified by MS/MS as the cleaved peptide. In comparison, much weaker signals were observed for the unprocessed N-terminal peptides that contained either HPG ($[M+H]^+$ *m/z* 1303.6) or N-formylated HPG ($[M+H]^+$ *m/z* 1331.6). Penultimate Ala enabled the almost complete removal of N-terminal HPG.

For IFNβ-2S, the cleaved N-terminal peptide produced the most abundant ion ($[M+H]^+$ *m/z* 1210.6). However, larger amounts of partially processed or unprocessed N-terminal peptides containing HPG ($[M+H]^+$ *m/z* 1319.6) or N-formylated HPG ($[M+H]^+$ *m/z* 1347.6) were detected for IFNβ-2S as compared to IFNβ-2A. Ser was less permissive than Ala with respect to processing of the N-terminal HPG.

The most surprising result was observed for IFNβ-2G. With Gly in the penultimate position, the signal intensities of the N-terminal peptides declined in the order: N-formylated HPG peptide > HPG peptide > cleaved peptide. IFNβ-2G expressed with HPG was very inefficiently processed in *E. coli*; this left a major portion of the protein in formylated and uncleaved forms. In comparison, analysis of IFNβ-2G expressed with Met showed that N-terminal Met was efficiently deformylated and cleaved as expected (Figure 3). Gly, being the smallest amino acid, supposedly fits into the MetAP substrate-binding pocket. With Gly at the penultimate position, complete cleavage of the N-terminal Met has been observed.[35,36] However, Gly does not always act as a good penultimate residue for MetAP cleavage.[33,34] For example, Met-Gly-Met was reported to be cleaved nearly four-times less efficiently by MetAP as compared to Met-Ala-Met.[34] The activity of MetAP, therefore, might be affected by other factors in addition to the side-chain length of the penultimate residue.[34,35] Our results show that Gly in the second position in IFNβ does not provide a good substrate for processing N-terminal HPG by either peptide deformylase or MetAP.

Based on our results with amino acids that favor retention of N-terminal AHA or HPG (Figure 2), we believe that small amounts of Met-containing species were produced during expression of IFNβ-2G, -2S, and -2A with HPG (or AHA). However, because N-terminal Met was cleaved efficiently by MetAP in these proteins, Met-containing species were not detected by MALDI-MS (Figure 4). Since the cleaved N-terminal peptide mass is the same for Met-peptide and UAA-peptide (Table 1), the cleaved peptide signal was composed of both sources. Thus, the percentage of cleavage of the N-terminal UAA might be over estimated by the small amount of Met incorporated into IFNβ.

Similar results were obtained for these three proteins expressed with AHA. That is, Ala was the most favorable penultimate residue with respect to promotion of deformylation and removal of N-terminal AHA (approaching 100% cleavage). With Ser or Gly at the penultimate position, incomplete processing was observed (Figure S1 in the Supporting Information). It appears that replacing Met with AHA or HPG at the N terminus decreases the activity of both the deformylase and MetAP to some extent, and the rules for processing N-terminal AHA or HPG deviate from what have been observed for the N-terminal Met. Specifically, whereas Ala is still a good substrate for deformylation and removal of N-terminal AHA or HPG in IFNβ, Gly is not. With Ser at the penultimate position, inefficient deformylation for fAHA-IFNβ and incomplete cleavage of N-terminal AHA and HPG were also observed.

N-terminal sequence (NTS) analysis

In addition to MALDI-MS, all recombinant proteins expressed with AHA or HPG were also analyzed by NTS. NTS analysis avoids the potential problems associated with variable ionization efficiencies of peptides in MALDI-MS experiments, and provides quantitative results on cleavage. However, NTS by Edman degradation alone is not sufficient to study the post-translational processing of proteins since it cannot detect proteins with N-terminal modifications, such as N-formylation or N-acetylation. Therefore, both analytical methods were used in the present studies.

Free AHA, DAB, and HPG were subjected to NTS analysis to establish their respective elution times in HPLC, and to examine their stability under NTS conditions. Under the experimental conditions used here, all three compounds separated well from other natural amino acids and from each other. The N-terminal protein sequences derived from NTS analysis correlated well with MALDI-MS data for all protein variants. The percentage of cleaved peptide was calculated based on the quantitation of cleaved peptide and the corresponding unprocessed peptide, and the results are summarized in Table 2.

Table 2. Percentage of N-terminal processed proteins based on NTS analysis. Based on repeated analysis for a few selected IFNβ, the standard deviation is approximately 5% for the assay.

Name	% cleaved product	
	with AHA	with HPG
IFNβ-2A	96	91
IFNβ-2S	80	80
IFNβ-2G	52	33
IFNβ-2H	8	0
IFNβ-2Q	0	0
IFNβ-2E	0	0

As shown in Table 2, N-terminal HPG and AHA are efficiently retained when Glu, Gln, and His occupy the penultimate position, although a small amount (~8%) of cleaved peptide was detected for IFNβ-2H with AHA. When the penultimate residue was Gly, 30–50% of cleaved peptides were observed. The actual percentage of cleavage was lower since the abundant f-HPG peptide detected by MS (Figure 4) was not detected in the NTS and thus not accounted for. When the penultimate residue was Ser, 80% of cleaved peptide was observed. When Ala resided at the penultimate position, the cleavage efficiency reached 90–100%. These results correlated extremely well with the MALDI-MS profiles. In fact, there are no significant differences between the percentages of cleavage based on MALDI-MS peak areas and those calculated from NTS.

For NTS analysis of proteins expressed with AHA, the first sequencing cycle was further characterized by converting the AHA peak area to its molar amount. This was achieved by analyzing a control synthetic AHA-containing peptide (X-SYNLLG, X = AHA) under the same conditions used for NTS analysis of test proteins. A correlation factor (ratio of AHA pmol to its peak area) was generated based on the synthetic peptide, and was used to calculate pmol of AHA detected in the test proteins (IFNβ-2E or IFNβ-2Q expressed with AHA). Such analysis showed that AHA occupies 85–100% of the first residue in these two proteins. The combined amounts of AHA and Met at the first cycle fully accounted for the total protein detected in subsequent cycles. Therefore, the DAB signal observed in the MALDI-MS is likely an artifact due to decomposition of AHA by laser irradiation.

Conclusions

Our results show that the processing of N-terminal AHA or HPG can be controlled by choosing an appropriate penultimate residue. We confirm previous observations that both AHA and HPG can be efficiently incorporated into recombinant proteins. These compounds act as efficient substrates for methionyl-tRNA synthetase, and yield expression levels equivalent to those obtained with Met. Of the three amino acids (Ala, Gly, and Ser) that favor removal of N-terminal Met, Ala is most efficient (90–100%) for promoting the removal of N-terminal AHA or HPG. Gly is not a good choice to remove N-terminal UAAs. On the other hand, it is likely that all penultimate residues that confer resistance to MetAP cleavage of N-terminal Met will also enable retention of N-terminal AHA or HPG, as demonstrated in this study for His, Gln, and Glu. Manipulating the expression level or substrate binding site of MetAP could provide other strategies for controlled processing of N-terminal UAAs. Based on this study, we were able to generate a modified human IFNβ that retains N-terminal AHA or HPG for further manipulation.

Experimental Section

Plasmid construction: An *E. coli* codon-optimized synthetic gene encoding a 20 kDa modified human IFNβ with only one Met codon (at the N terminus) was amplified by PCR by using overlapping oligonucleotides (Operon). The synthetic gene was cloned into the pQE30 expression vector (Qiagen) under control of a T5-lac-lac promoter/regulator by using standard methods. Different penultimate residues were introduced into IFNβ by site-directed mutagenesis. The oligonucleotide sequences encoding the corresponding N-terminal tryptic peptides are listed in Table 1. The sequences of all cloned genes were confirmed by DNA sequencing.

Incorporation of AHA or HPG into recombinant protein: The pQE30 expression vector containing the synthetic IFNβ gene was transformed into Met auxotrophic *E. coli* strain M15MA[39] with helper plasmid pREP4 (Qiagen). Dual antibiotics (100 mg L^{-1} carbenicillin and 50 mg L^{-1} kanamycin) were used in all culture media for selection of both pQE30 and pREP4 plasmids. A single colony was used to inoculate Luria–Bertani (LB) broth for overnight growth at 37 °C. The overnight culture was diluted 50-fold the next morning into fresh LB media, and cells were allowed to grow at 37 °C until the OD_{600} reached approximately 1. The culture was centrifuged to remove LB media. Cells were resuspended in M9 minimal media and grown at 37 °C for 30 min. Cells were centrifuged again, and then resuspended in M9 minimal media supplemented with all 19 amino acids except Met (Arg, His, and Trp: 15 mg L^{-1}; Ile, Lys, and Tyr: 23 mg L^{-1}; Phe: 38 mg L^{-1}; Leu: 75 mg L^{-1}; Thr: 150 mg L^{-1}; Val: 113 mg L^{-1}; all others: 19 mg L^{-1}). The cell culture was further supplemented with AHA or HPG (50 mg L^{-1}; Medchem, WA, USA). Parallel cultures with and without Met (25 mg L^{-1}) were included as controls. A final concentration of 1 mM IPTG was added to induce recombinant protein expression, and cells were harvested 2 h postinduction.

Analysis of recombinant proteins by matrix assisted laser desorption/ionization mass spectrometry (MALDI-MS): Recombinant IFNβ was separated from most of the endogenous *E. coli* proteins by SDS-PAGE (4–20%) under reducing conditions. The IFNβ band visualized by Coomassie blue stain or SureBlue Safestain (Invitrogen) was cut out from the gel and subjected to overnight trypsin digestion at 37 °C after destaining and modification with iodoacetamide. The sample was completely dried, and redissolved in trifluoroacetic acid (TFA; 0.1%) containing acetonitrile (2%). The sample was desalted by using wall-coated C18 micropipette tips (New Objective, Woburn, MA, USA), and eluted with acetonitrile (60%) with 0.1% TFA (10–20 μL). The eluted sample was mixed with an equal volume of α-cyano-4-hydroxycinnamic acid (10 mg mL^{-1}) in acetonitrile (70%) containing TFA (0.1%) and ammonium dihydrogen phosphate (5 mM; Aldrich). An aliquot (1 μL) was then spotted on an Opti-TOF 96-well insert (Applied Biosystems, Framingham, MA, USA) and analyzed by using a 4800 MALDI TOF/TOF analyzer (Applied Biosystems) that was calibrated for a mass range of 900 to 4000 Da with "4700 calibration mix" (Applied Biosystems). For MS data acquisition, 100 laser shots were fired at each of 20 different random locations on the sample spot (total of 2000 laser shots per sample). For tandem mass spectrometry (MS/MS) data acquisition, up to 3000 laser shots were accumulated per precursor ion. The N-terminal peptide assignments were confirmed by the presence of anticipated fragment ions in their respective tandem mass spectra.

Analysis of recombinant proteins by N-terminal sequencing (NTS): Recombinant IFNβ was expressed as inclusion bodies in *E. coli* and purified by detergent washes of cell pellets or by organic extraction and acid precipitation.[50] The IFNβ protein was further separated from contaminating *E. coli* proteins with SDS-PAGE (4–20%). After being transferred to a PVDF membrane, the IFNβ band was cut out and analyzed with five cycles of Edman degradation on an Applied Biosystems protein sequencer equipped with an online HPLC system (Keck Laboratory, Yale University). Routinely, phenylthiohydantoin (PTH) standards (1.0 pmol) were used for calibration. In addition, the S4 solvent that transfers the PTH derivatives to HPLC contained PTH-norvaline, which acts as an internal calibrant to independently monitor transfer to the HPLC. Free amino acid analogues (HPG, AHA, and diaminobutyric acid (DAB)) were subjected to NTS analysis to establish their elution times and stability under sequencing conditions. A synthetic peptide containing AHA at the N terminus (X-SYNLLG, where X is AHA, designated AHA-SYNLLG, custom synthesized by Medchem, Federal Way, WA, USA) was used as a standard to convert the AHA peak area to its molar amount. The percentage of cleaved product was calculated by dividing the amount of protein initiated at the second position by the sum of the amounts of protein initiated at both the first and second positions, as described by Liao et al.[35]

Acknowledgements

We would like to thank Inchan Kwon, Tom Graddis, Kurt Shanebeck, Scott Bloom, and J. Myron Crawford for their advice and assistance. Research on UAAs at Caltech is supported by NIH Grant GM62523.

Keywords: methionine aminopeptidase · N-terminal processing · protein engineering · protein modifications · unnatural amino acids

[1] K. Forchhammer, W. Leinfelder, A. Bock, *Nature* **1989**, *342*, 453–456.
[2] F. Zinoni, A. Birkmann, W. Leinfelder, A. Bock, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 3156–3160.
[3] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459–1462.
[4] C. Polycarpo, A. Ambrogelly, A. Berube, S. M. Winbush, J. A. McCloskey, P. F. Crain, J. L. Wood, D. Soll, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12450–12454.
[5] N. Budisa, *Engineering the Genetic Code*, Wiley-VCH, Weinheim, **2006**.
[6] N. C. Yoder, K. Kumar, *Chem. Soc. Rev.* **2002**, *31*, 335–341.
[7] Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado, D. A. Tirrell, *Angew. Chem.* **2001**, *113*, 1542–1544; *Angew. Chem. Int. Ed.* **2001**, *40*, 1494–1496.
[8] S. Son, I. C. Tanrikulu, D. A. Tirrell, *ChemBioChem* **2006**, *7*, 1251–1257.
[9] S. K. Wright, M. M. Kish, R. E. Viola, *J. Biol. Chem.* **2000**, *275*, 31689–31694.
[10] J. E. Oh, K. H. Lee, *Bioorg. Med. Chem.* **1999**, *7*, 2985–2990.
[11] S. H. W. Beiboer, B. van den Berg, N. Dekker, R. C. Cox, H. M. Verheij, *Protein Eng.* **1996**, *9*, 345–352.
[12] N. Muranaka, T. Hohsaka, M. Sisido, *FEBS Lett.* **2002**, *510*, 10–12.
[13] M. Taki, T. Hohsaka, H. Murakami, K. Taira, M. Sisido, *FEBS Lett.* **2001**, *507*, 35–38.
[14] M. Grace, S. Youngster, G. Gitlin, W. Sydor, L. Xie, L. Westreich, S. Jacobs, D. Brassard, J. Bausch, R. Bordens, *J. Interferon Cytokine Res.* **2001**, *21*, 1103–1115.
[15] A. Deiters, T. A. Cropp, D. Summerer, M. Mukherji, P. G. Schultz, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5743–5745.
[16] T. Hohsaka, M. Sisido, *Curr. Opin. Chem. Biol.* **2002**, *6*, 809–815.
[17] J. M. Bacher, A. D. Ellington, *Methods Mol. Biol.* **2007**, *352*, 23–34.
[18] E. A. Rodriguez, H. A. Lester, D. A. Dougherty, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8650–8655.
[19] D. R. Liu, T. J. Magliery, M. Pastrnak, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10092–10097.
[20] A. J. Link, D. A. Tirrell, *Methods* **2005**, *36*, 291–298.
[21] R. Furter, *Protein Sci.* **1998**, *7*, 419–426.
[22] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498–500.
[23] J. Xie, P. G. Schultz, *Methods* **2005**, *36*, 227–238.
[24] K. L. Kiick, R. Weberskirch, D. A. Tirrell, *FEBS Lett.* **2001**, *502*, 25–30.
[25] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19–24.
[26] J. C. van Hest, D. A. Tirrell, *FEBS Lett.* **1998**, *428*, 68–70.
[27] K. L. Kiick, J. C. M. van Hest, D. A. Tirrell, *Angew. Chem.* **2000**, *112*, 2232–2236; *Angew. Chem. Int. Ed.* **2000**, *39*, 2148–2152.
[28] C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057–3064.
[29] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem.* **2002**, *114*, 2708–2711; *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.
[30] S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony, B. G. Davis, *Nature* **2007**, *446*, 1105–1109.
[31] P. T. Rajagopalan, A. Datta, D. Pei, *Biochemistry* **1997**, *36*, 13910–13918.
[32] T. Meinnel, S. Blanquet, *J. Bacteriol.* **1995**, *177*, 1883–1887.
[33] A. J. Link, K. Robison, G. M. Church, *Electrophoresis* **1997**, *18*, 1259–1313.
[34] A. Ben-Bassat, K. Bauer, S. Y. Chang, K. Myambo, A. Boosman, S. Chang, *J. Bacteriol.* **1987**, *169*, 751–757.
[35] Y. D. Liao, J. C. Jeng, C. F. Wang, S. C. Wang, S. T. Chang, *Protein Sci.* **2004**, *13*, 1802–1810.
[36] P. H. Hirel, M. J. Schmitter, P. Dessen, G. Fayat, S. Blanquet, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 8247–8251.
[37] F. Sherman, J. W. Stewart, S. Tsunasawa, *Bioessays* **1985**, *3*, 27–31.
[38] N. Budisa, O. Pipitone, I. Siwanowicz, M. Rubini, P. P. Pal, T. A. Holak, M. L. Gelmi, *Chem. Biodiversity* **2004**, *1*, 1465–1475.
[39] A. J. Link, D. A. Tirrell, *J. Am. Chem. Soc.* **2003**, *125*, 11164–11165.
[40] J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 1282–1288.
[41] W. T. Lowther, A. M. Orville, D. T. Madden, S. Lim, D. H. Rich, B. W. Matthews, *Biochemistry* **1999**, *38*, 7678–7688.
[42] W. T. Lowther, B. W. Matthews, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **2000**, *1477*, 157–167.
[43] J. Tang, G. Hernandez, D. M. LeMaster, *Protein Expression Purif.* **2004**, *36*, 100–105.
[44] W. C. Warren, K. A. Bentle, M. R. Schlittler, A. C. Schwane, J. P. O'Neil, G. Bogosian, *Gene* **1996**, *174*, 235–238.
[45] F. A. Carey, R. J. Sundberg, *Advanced Organic Chemistry Part B: Reactions and Synthesis* Springer, New York, **2001**, pp. 642–644.
[46] K. Kirshenbaum, I. S. Carrico, D. A. Tirrell, *ChemBioChem* **2002**, *3*, 235–237.
[47] J. M. Guillon, T. Meinnel, Y. Mechulam, C. Lazennec, S. Blanquet, G. Fayat, *J. Mol. Biol.* **1992**, *224*, 359–367.
[48] C. P. Lee, B. L. Seong, U. L. RajBhandary, *J. Biol. Chem.* **1991**, *266*, 18012–18017.
[49] E. Schmitt, M. Panvert, S. Blanquet, Y. Mechulam, *EMBO J.* **1998**, *17*, 6819–6826.
[50] L. S. Lin, R. Yamamoto, R. J. Drummond, *Methods Enzymol.* **1986**, *119*, 183–192.

Received: July 10, 2007
Published online on December 20, 2007

